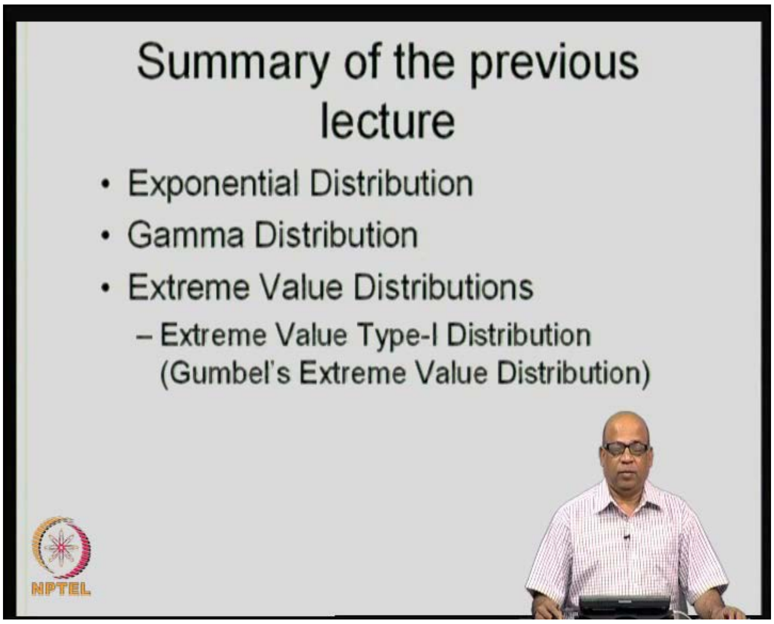
Stochastic Hydrology

Prof. P. P. Mujumdar
Department of Civil Engineering
Indian Institute of Science, Bangalore

Lecture No. # 07

Parameter Estimation

(Refer Slide Time: 00:22)



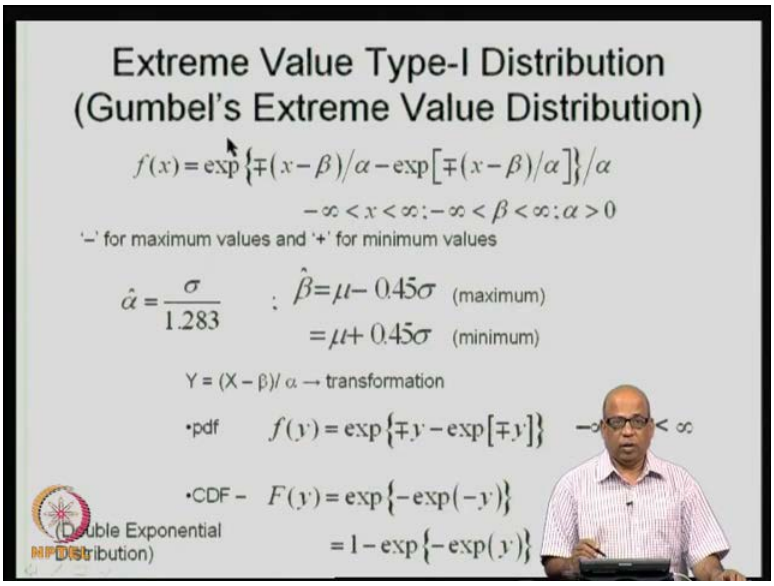
Good morning and welcome to this seventh lecture of the course stochastic hydrology. In the last lecture, you can recall now that we discussed the exponential distribution, the gamma distribution, and we introduce the extreme value distributions. If you recall the exponential distribution is generally used when we are interested in the time between two critical events. Let us say the time between two flood events, which it is self is a random variable, then we would be generally using the exponential distribution. The gamma distribution is in fact a family of distributions, and you recall that exponential distribution is in fact a special case of gamma distribution, where we put eta is equal to 1 in the gamma distribution pdf, to obtain the pdf of exponential distribution.

The gamma distribution is a typically used for rainfall distributions across several time frames or in many cases we use the gamma distribution also for a seasonal rainfall monthly rainfall and so on. Then we went on to discuss the extreme value distribution, if you recall we introduced three types of extreme value distribution, they are called as type I, type II and type III extreme value distributions. The extreme value distributions are very commonly used in hydrology, because we would be interested often in the peak flows or the minimum flows, the high intensity rainfall. For example, the maximum intensity, daily intensity of the rainfall, and also the low flows when we are interested in water quality or drought phenomena, and so on.

Now, the extreme value type I distribution is a distribution, which draws from a parent distribution, which is unbounded in the direction of the extreme values that we are interested in for example, if you are interested in the maximum values, then you should be unbounded on the positive side. If you are interested in the minimum values it should be unbounded on the lower side. So, extreme value type I distribution essentially comes from the parent distribution, which are unbounded in the direction of the extreme values that we are interested in and we introduce the Gumbel’s extreme value distributions. We will review what we discussed in the last class in the first two slides of today class; the extreme value type II distribution is not very commonly used in hydrology.

So, we then go on to discuss the extreme value type III distribution, which is derived from the parent distribution, which is bounded in the direction of the extreme value that we are interested in for example, if you are interested in the minimum values it should be bounded on the lower side. We will introduce in today class the extreme value type III distribution. We discuss the extreme value type I distribution in the last class, which is also called the Gumbel’s extreme value distribution. So, commonly in the hydrology literature, we find this commentator that is Gumbel’s extreme values distribution or simply Gumbel’s distribution. We introduce this pdf in the last class, f of x given by the expression, which has two parameters alpha and beta, and we have the relationships for alpha and beta, which are estimated from the samples using the standard deviation and the mean.

(Refer Slide Time: 03:56)

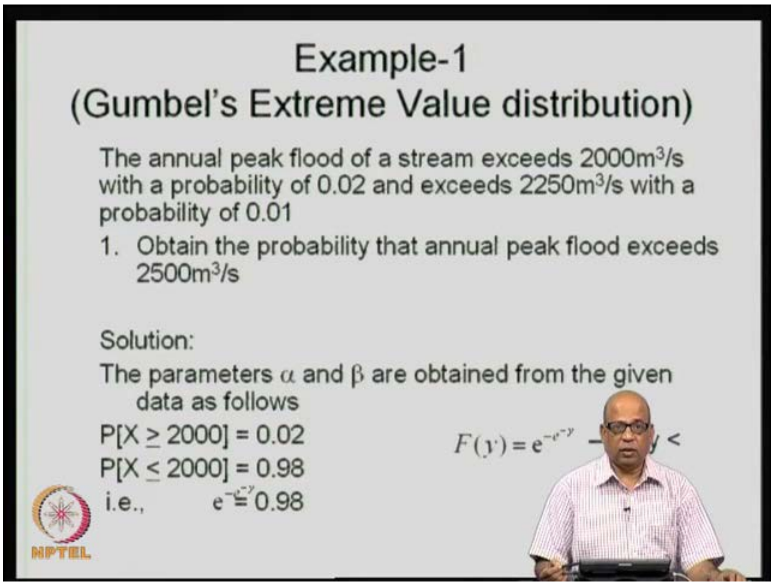


So, if you are looking at the extreme value distribution type I for maximum values, then we use the negative sign here in the pdf, and for minimum values we use the positive sign in the pdf correspondingly the parameter beta. The estimate for the parameter beta, we use the negative sign here it turns out to be mu minus 0.45 sigma for the maximum values and for the minimum values you get mu plus 0.45 sigma. So, given the sample values, the sample estimate for the standard deviation and you also get the sample estimate for the mean from which you can estimate alpha and beta.

We used the transformation y is equal to x minus beta or alpha which simplifies the pdf, and we write the pdf for the transformed variable y as follows F of y is equal to exponential minus plus y minus exponentials minus plus y and again minus is used for minimum values like in the case of our earlier expression and plus is used for the maximum values. The other way round we use the maximum we use the negative value for the maximum value and positive values for the minimum values.

So, when we take the cdf, we consider the cdf, you get f of y is equal to exponential minus exponential minus y. So, this is a convenient form to use for maximum values and for the minimum values it comes out to be 1 minus exponential minus exponential y for the minimum values. So, as you can see, it can be written as e to the power minus e to the power minus y. So, this is called as the double exponential distribution. So, when we use the transformation y is equal to x minus beta over alpha, the cdf comes out to be in an elegant form e to the power minus e to the power minus y for the maximum values, let us consider an example now for the Gumbel’s extreme value distribution. If you are looking at the positive extreme value which means the peak floods or the peak run off and so on, we will take this example which demonstrates the positive extreme value.

(Refer Slide time: 06:32)



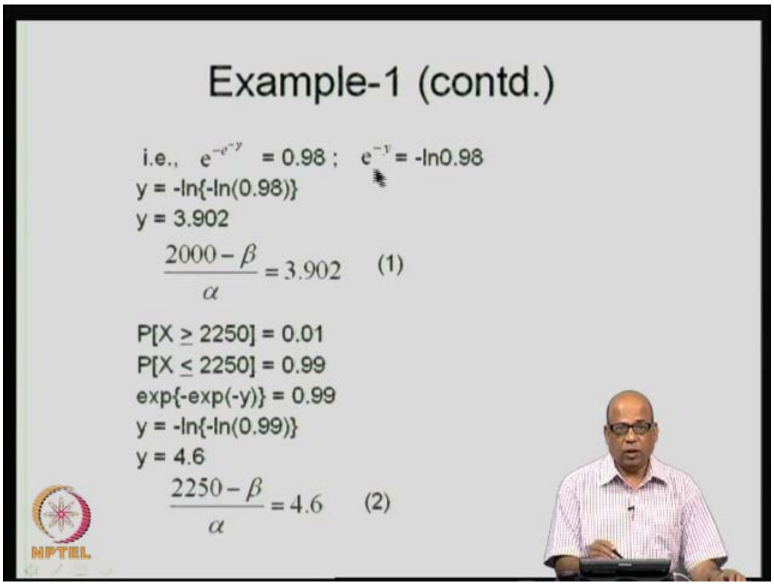
The annual peak flood of a stream it is own that it exceeds a value of 2000 meter cube per second with the probability of 0.02 and it exceed a value of 2250 meter cube per second with the probability of 0.01. So, these are known from the samples and then we need to obtain the probability that annual peak flood exceeds 2500 meter cube per second assume of course that the peak flood follows a Gumbel’s extreme value distribution. So, how do we do this? So, like we did with all earlier distributions, we first estimate the parameter alpha and beta to estimate the parameter. We consider the fact that probability of x being greater than equal to 2000 is given to be 0.02, from which you write probability of x being less than equal to 2000 as 0.98 and you know that f of y, which is the cdf is e to the power minus e to the power of minus y.

If you use the transformation, write e to the power minus e to the power of minus y is equal to 0.98 for that particular value of y. We write this as e to the power minus e to the power of minus y is equal to 0.98 and therefore, e to the power of minus y by taking logarithms turns out to be minus log 0.98 from which we write y is equal to minus log of this particular term, which is minus log 0.98 from which you get y is equal to 3.902.

What is y? You recall that y is a transformation like this the y is equal to X minus beta over alpha. So, this is the y value which is 2000, which is the particular X value that we are talking about minus beta or alpha. So, that should be equal to y, which is 3.902; this is our first condition that we generate.

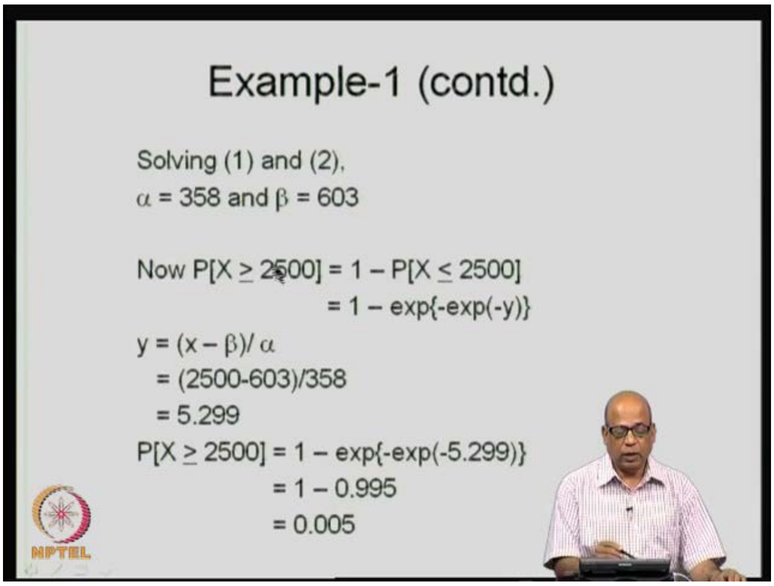
Then from the second condition, which is given as probability that the peak flows exceed 2250 meter cube per second is 0.01. So, we write probability that X is greater than equal to 2250 must be equal to 0.01 from which we write probability of X being less than equal to that particular value will be equal to 1 minus of this is equal to 0.99 again, we use the same method and the estimate y. Y in this case transfer to be 4.6, which is y as you know is a transformation X minus beta or alpha. So, 2250 which is particular value of X minus beta over alpha this is equal to 4.6. So, we generated two conditions we solve these two equations to get the two unknown alpha and beta.

(Refer Slide Time: 08:19)



So, we get alpha is equal to 358 and beta is equal to 603. So, like in any earlier distributions the first step is always to estimate the parameters from the sample values. So, estimated here in this particular case alpha is equal to 358 and beta is equal to 603 now, we are ask the question probability, what is the probability that x will be greater than equal to 2500 meter cube per second. So, this we write it as equal to 1 minus probability of x being less than equal to 2500 which from the expression for f of y, we write it as 1 minus exponential minus exponential minus 5 that is 1 minus e to the power minus e to the power minus 5, where y is equal to x minus beta or alpha and x is that particular value which in this case is 2500 beta we have estimated it to be 603 and alpha we have estimated it to be 358. So, we get the corresponding value of y as 5.299. So, we write probability of x being less greater than equal to 2500 as equal to 1 minus exponential minus exponential minus y, which is 1 minus exponential minus exponential minus y, which is 5.299.

(Refer Slide Time: 10:02)

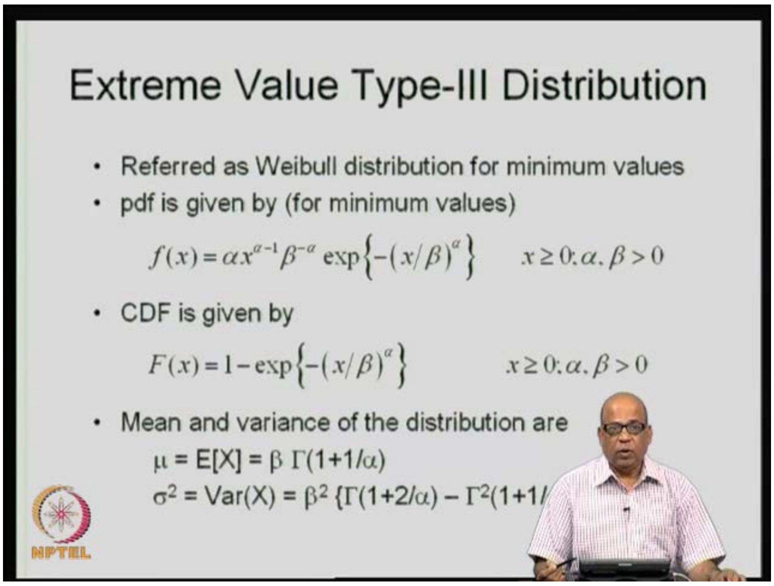


So, this turns out to be 0.005, which is probability that x being greater than equal to 2500 will be 0.005 which is fairly low probability. Then we move on to what we call as the type III distribution. If you recall the type III distribution is bounded on the side of the extreme value that we are interested in for example, if you are interested in the minimum values then it should be bounded on the lower side, if you are interested in the maximum values then you should be bounded on the higher side, unlike the type I distribution which will be unbounded, the parent distribution will be unbounded in the case of distribution on the side of the extreme.

So, the extreme value type III is bounded it is derived from a parent distribution which is bounded on the side of the extreme in which we are interested in now, for the minimum value, if you are considering the type III distribution for the minimum values, this is referring to as the Weibull distribution; whenever we say it is a Weibull distribution, we are talking about the extreme value type III distribution. When we are considering the distribution for the minimum value the pdf is given by f of x is equal to alpha x to the power of alpha minus 1 beta to the power minus alpha exponential of minus x by beta to the power alpha.

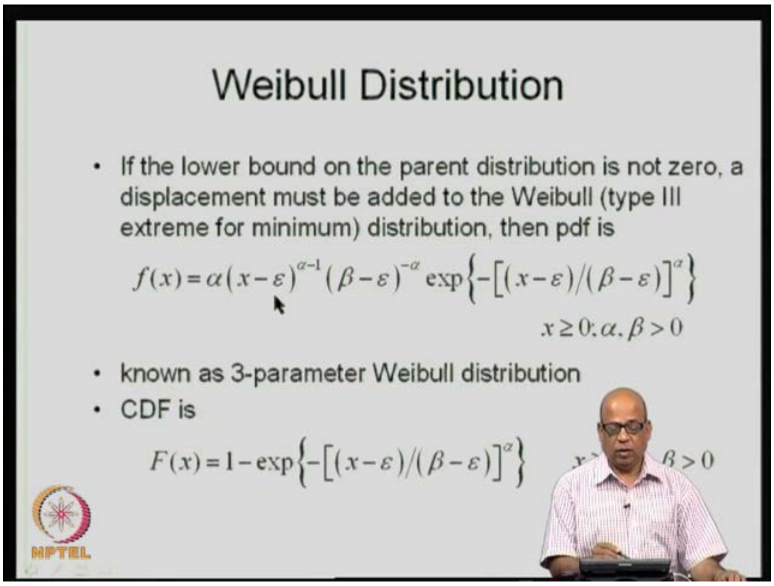
So, this has two parameters alpha and beta, and this is defined for x non negative x greater than equal to 0 and alpha and beta both positive from this you can derive the cdf, cumulated distribution function as f of x is equal to 1 minus exponential minus x by beta to the power alpha. Now, it can be shown that the mean and the variance of this distribution defined as the given here, will be the means is which is the expected value of x, is beta into gamma function of 1 plus 1 over alpha. Similarly, the variance sigma square is given by beta square into brackets the gamma functions of 1 plus 2 by alpha minus the gamma function the square of 1 plus 1 by alpha. So, given the sample values we should be able to estimate alpha and beta by using these two moments, mu and sigma square.

(Refer Slide Time: 12:08)



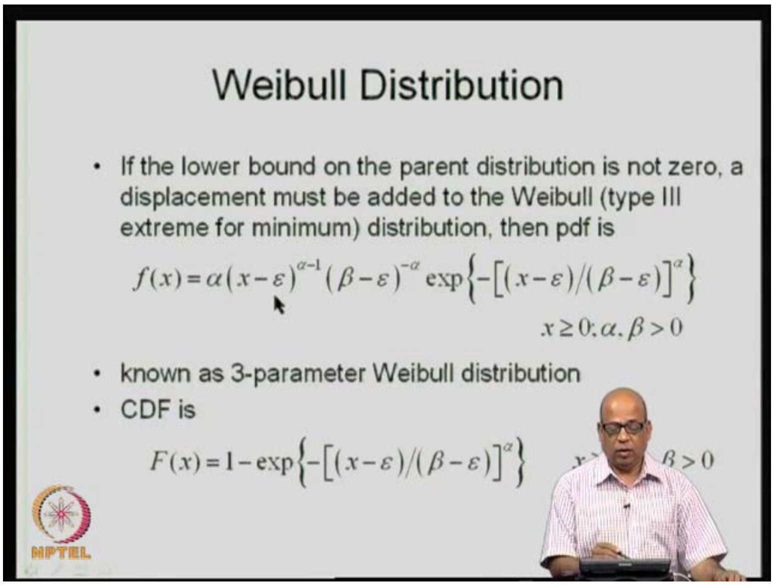
Now, if the low bound on the parent distribution is not 0, then we put a displacement epsilon. So, we add a displacement and then redefined our pdf. So, whenever we do not have the lower bound to be 0 then we add corresponding to the displacement that the parent distribution has and this is known as three parameter Weibull distribution and associated with this, we have the cdf as given by this expression f of x is equal to 1 minus exponential minus x minus alpha. I am sorry x minus epsilon divided by beta minus epsilon raise to the power alpha just compare this with our earlier cdf which where we had 1 minus exponential minus of x by beta. So, all we have done is we have subtracted a displacement factor, a displacement parameter x minus epsilon here and beta minus epsilon here that is all the difference and this is called as the three parameter Weibull distribution, because it has an addition parameter of epsilon apart from two original parameter alpha and beta.

(Refer Slide Time: 15:37)



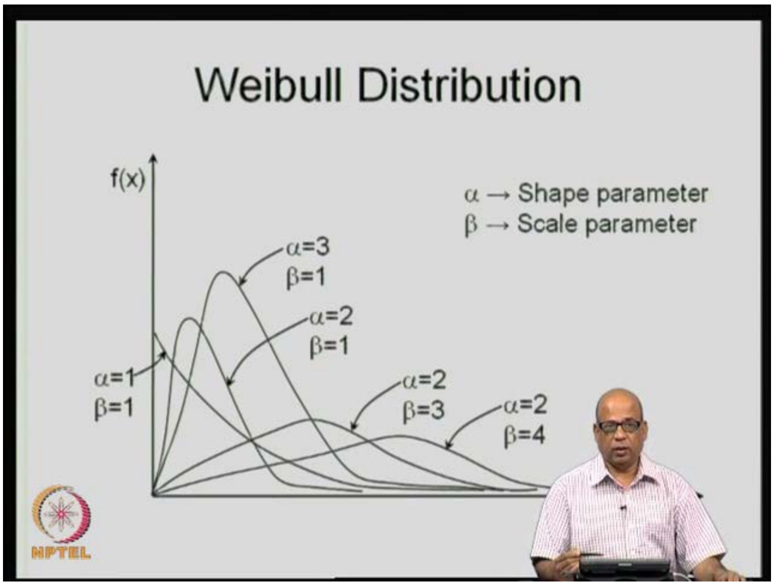
Now, in this case again the moments that is the first moment mu is given by epsilon plus beta minus epsilon gamma function 1 plus 1 over alpha and the variants, which is the seconds moments about the means is given by beta minus epsilon square and gamma function of 1 plus 2 by alpha minus the gamma function the square of 1 plus 1 by alpha. So, again from the sample estimates of mu and sigma square, you should be able to get the parameters alpha and beta by using also the third moment in that particular case. If you have three parameters, we will discuss the parameter estimation hopefully in today class of different distributions, the methods of different parameter distributions.

(Refer Slide Time: 16:10)



So, Weibull distribution with the two parameter alpha and beta appear something like this as we discussed in the case of gamma distribution. The alpha parameter determines the shape of distribution for example, for a given beta, beta is equal to 1, if we change the alpha the shape will change alpha is equal to 1 provides this particular shape for beta is equal to 1, if you change the alpha, it provides this particular shape. So, alpha governs the shape of a parameter, shape of the distribution and beta governs the scale of the parameter for example, alpha is equal to 2 here, beta is equal to 1gives you this scale and beta is equal to 3 gives you this scale. So, alpha is therefore, called as a shape parameter and beta is called as the scale parameter let us consider an example we have remembered when we are talking about the minimum values that are the extreme distributions for the minimum values.

(Refer Slide Time: 17:38)

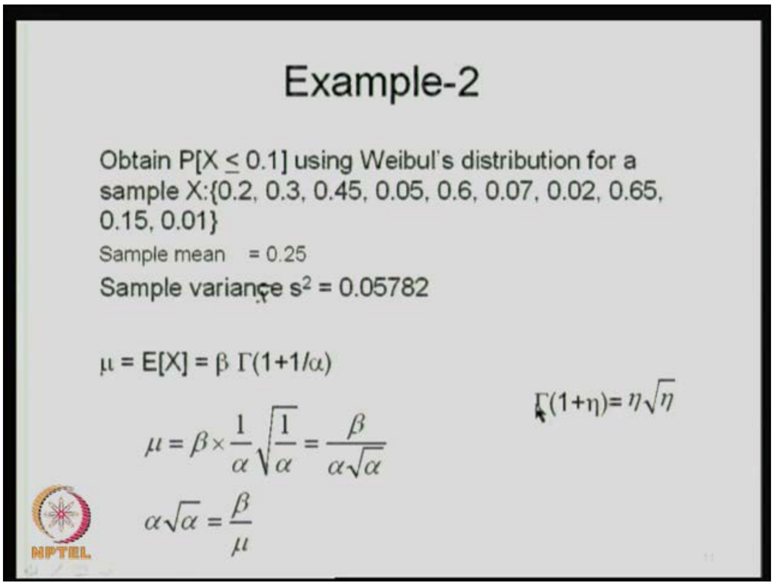


We call it as Weibull's distribution. So, Weibull's distribution, we are always talking about the lower extreme. So, we will obtain the probability of x being less than 0.1 using Weibull distribution for a sample x is equal to as given here. Now, these are the low flows in a stream and we are interested in the distribution of the low flows and where interested in the probability that x is less than according to 0.1 as I mentioned earlier interest of on exist in low flows. When we are talking about water quality or when we are talking about drought situation where the water availability of concern and therefore, we will be interested in probability of flow being less than equal to a certain value 0.1. So, these are the low flows maybe the minimum annual flows and so on.

Now, when we take the mean of this the sample mean is 0.25 this is simply the mean or the estimate of the mean here, that is the arithmetic average x bar by n and the sample variance, s square is 0.05782. Now, we use these 2 and equate them to the mean as obtain for the Weibull distribution and the variance as obtain for the Weibull distribution. So, this is in fact, called as methods of movement which will introduce into this class subsequently. So, essentially what we are doing is the means as obtained from the sample or estimated from the sample and the variance as estimated from the sample are equated to the mean as obtained from the theoretical distribution and the variance as obtained from the theoretical distribution and thus we generate two equation and solve for the two unknowns alpha and beta.

So, mean as obtain from the theoretical distribution as was given here earlier for the two parameter of Weibull's distribution, the mean is given by beta into gamma function of 1 plus 1 by alpha and the variance is given by sigma square is equal to beta square into brackets the gamma functions 1 plus 2 by alpha minus gamma square 1 plus 1 by alpha. So, these are the expressions that we use and then obtain the two conditions mu is equal to expected value of x is equal to beta into gamma function of 1 plus 1 by alpha which is equal to we write mu is equal to beta into this one plus one by alpha the gamma function. We call that when introduce the gamma function in the gamma distribution we said gamma function of 1 plus eta can be written as eta as root eta.

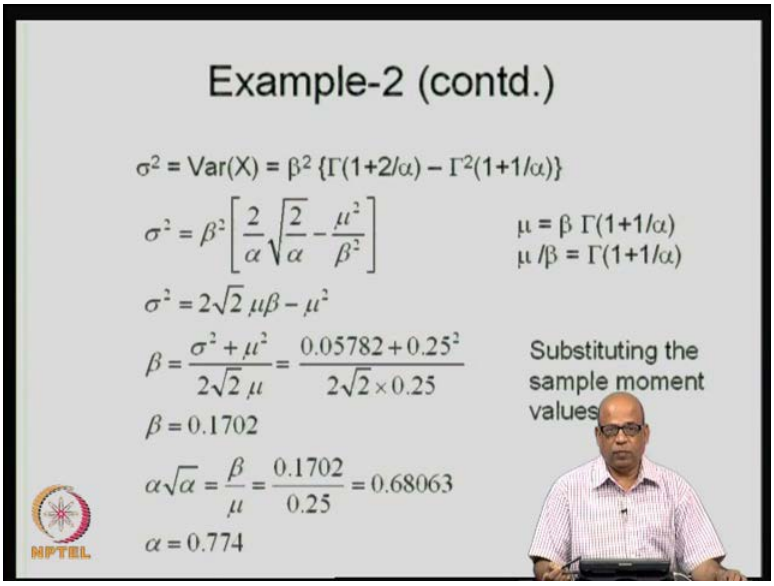
(Refer Slide Time: 20:38)



So, that is the relationship that we use here and then write it as beta into 1 by alpha root 1 by alpha which turns out to be beta divided by alpha root alpha or we write it as alpha root alpha is equal to beta by mu from this expression. Then we also use the variance sigma square for the Weibull distribution sigma square is given by beta square into the gamma function 1 plus 2 by alpha minus gamma square 1 plus 1 by alpha again, this 1 plus 2 by alpha we write it as 2 by alpha root 2 by alpha. So, we had taken out beta square and then this is the first term gamma function of 1 by 1 plus 2 by alpha we write it as 2 by alpha root 2 by alpha minus the gamma function square 1 plus 1 by alpha we call that we wrote mu is equal to beta into gamma function of 1 plus 1 by alpha.

Mu by beta we write it as gamma function of 1 plus 1 by alpha and therefore, the gamma square 1 plus 1 by alpha we write it as from here mu square by beta square and from there you get sigma square is equal to 2 root 2 mu b this is what we are writing now, 2 root 2 mu beta minus mu square when you simplify this and then we get beta is equal to 0.1702 when you substitute the value for mu as well as sigma square. This is mu square and this is sigma square. So, once you get beta you go to alpha root alpha is equal to beta by mu and get alpha is equal to 0.774. So, this is how we are obtaining the two parameters.

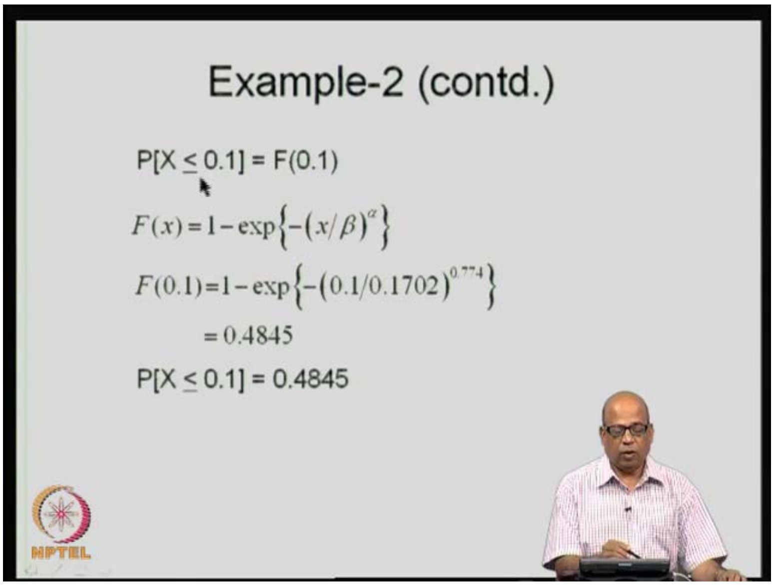
(Refer Slide Time: 23:15)



We essentially equated the first moment as estimated from the sample which is the arithmetic average to the first moment as obtained from your theoretical Weibull distribution and, similarly the second moment about the mean which is the variance we get the sample estimate and equated it to the second moment about the mean which is the variance as defined for your Weibull distribution theoretical Weibull distribution and thus to generate two equations and obtain the parameter alpha and beta. This is in fact, called as the method of moments for estimation of the parameter. This we would be introducing in today’s class subsequently. So, once you get alpha and beta the Weibull distribution is completely defined and therefore, we can talk about the probabilities associated with the random variable following the Weibull’s distribution.

So, we write probability of x being less than equal to 0.1 which is by definition equal to f of 0.1 F of 0.1 and f of x for the Weibull distribution is by 1 minus exponential minus x by beta to the power half this x is that particular value of x. That we are interested in that which is 0.1 in this case and beta is estimated here as 0.1702 and alpha is estimated here as 0.774. So, we substitute these values and get f of 0.1 is equal to 0.4845 or probability of x being less than equal to 0.1 is 0.4845. So, essentially in the extreme value distributions what we do is that we use either the type one distribution or the type III distribution depending on whether you are interested in the lower extreme or the higher extreme.

(Refer Slide Time: 25:00)



Now, the type I distribution essentially comes from the parents distributions which are unbounded on the side of the extreme that you are looking for example, if you are looking for the minimum values then the parents we which has a lower bounded or the lower extreme unbounded for example, the normal distribution which has the lower extreme unbounded, if you are interested in type I distribution which and you want to use the maximum values or the hear extremes parent distributions should have a higher bound which is the bound on the right side should be unbounded. Again the normal distribution is an example or the log normal distribution is an example, gamma distribution is an example, which have the higher bound as unbounded we use the type III distributions when we are looking or let say minimum values and the parent distribution has the lower bound the lower bounded. So, in the type I distribution, type

III distribution we are talking about the parent distribution being bounded on the side of the stream that we are looking for. When we are using the type III distributions for the minimum values, it is called as the Weibull distribution which is more commonly used in hydrology for events like low stream flows, low rainfalls, minimum temperatures and so on.

When we are looking at the minimum value or the lower extreme values then we use the Weibull distribution. So, essentially or the in hydrology most commonly we use the type I distribution for the maximum values typically the floods peaks and high stream flows high intensities of rainfall and such higher values of the event. Higher values of particular random variable we will be using the type I distribution which is called as a Gumbel’s distribution and type III distribution we typically use in hydrology for the lower bound of the random variables, typically for low stream flows, minimum rainfall, minimum temperature and such events and this is also called as the Weibull distribution.

So, we have covered a range of distribution we started the normal distribution then we went on to the log normal distribution the normal distribution is most commonly used in for normal processes for example, average stream flows, average annual stream flows, and seasonal rainfall where you are accumulating large periods of time. Such normal events we use normal distribution it is most commonly used in hydrology, but the limitation of the normal distribution, if you recall are limitations are that there is a finite probability associated with negative values and often in hydrology we deal with non negative value non negative variable for example, stream rainfall of the negative and so on. Also the normal distribution is a perfectly symmetrical distribution and in hydrology we often get situation where the distributions are not symmetrical and they are often skewed distributions. It is in these situations that the log normal distribution has advantage that it is defined for non negative values of x. So, we graduated on to the log normal distribution where we are talking about a positive skewed distributions and it is defined essentially for non negative values of the argument x.

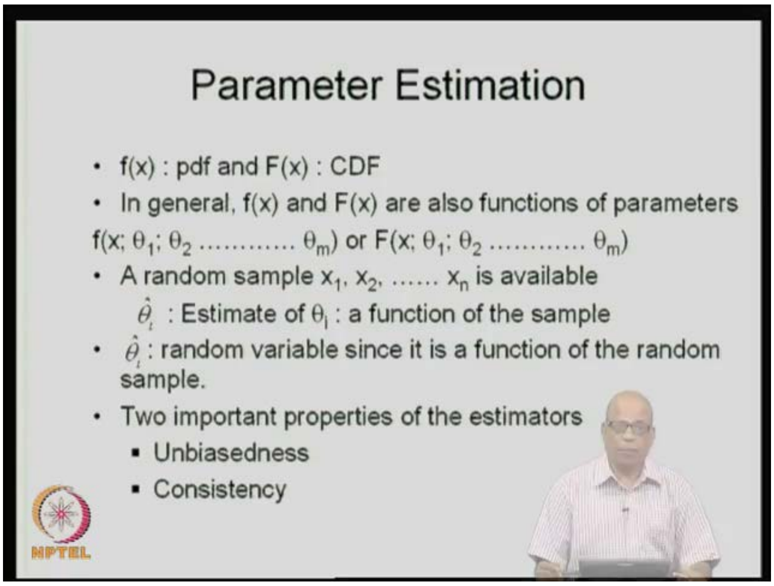
Then we also went on to the exponential distribution which is typically used for time between two critical events and the we introduce the gamma distribution, which is the family of distributions in of which exponential is a specific case then we went on to introduce a extreme value distribution when we are interested in really high value or low value of a particular random variable. Now, these are some of the distributions that we commonly used, but they are not as set of distributions as any function which satisfies the two conditions namely f of x is non negative and the integral between minus infinity to plus infinity with respect to x must be equal to 1, any function that satisfies these two conditions that is a potential probability density function. So, there are virtually infinitely many such probability distributions.

We have, in this particular course introduced those distributions which are commonly used in hydrology at first level. Now we move on to another topic where we are interested in estimating the parameters of a distribution. As we know any distribution which satisfies those two conditions I just mentioned is a potential a distribution and therefore, we have virtually infinitely many such distributions each distributions will have certain parameter for example, the normal distribution had two parameters and the Weibull distribution that we just introduce has had two parameters alpha and beta.

The exponential distribution had a parameter lambda the gamma distribution had two parameters log normal distribution had two parameters and so on. So, typically a distribution will have one or more than one parameters given a sample we should we should know how to estimate the parameters from the sample parameters of the distribution from the sample for example, you have a sample of observed stream flows for the last 50 years. How do we make use of this sample of observed stream flows to estimate the parameters of let say normal distribution assuming that this sample follows a normal distribution or if it follows a let us say exponential distribution, if you we have a sample of time between two critical events and we have collected that sample over a past historical period.

How do we estimate the parameters of the exponential distribution, assuming that the sample follows the exponential distribution? So, we now introduce the important topic of parameter estimation. So, essentially the problem that we will be concern within parameter estimation is that you have a sample of observe values and from these samples we want to estimate the parameters of a particular distribution. So, how do we go about this? First we introduce the concept of the parameter estimation itself.

(Refer Slide Time: 32:34)



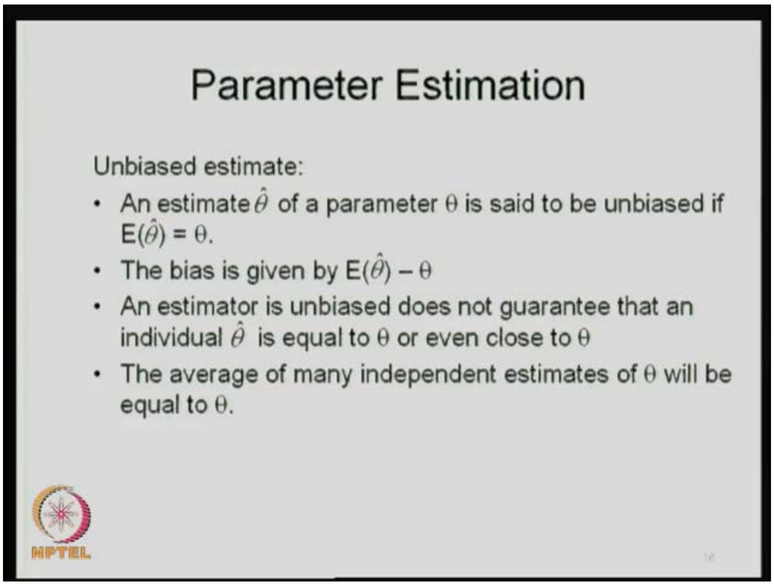
We have been talking about the pdf and the cdf, we denoted the pdf by f of x small f of x and the cdf by the F of X, but in general f of x and f of x the small f of x and the capital F of X are also functions of the parameter and therefore, we should be writing f of x semicolon theta 1 theta 2 etcetera, theta m. So, this has m parameters, similarly f of x has m parameters theta 1, theta 2, etcetera, theta m now, we want to estimate these parameters theta 1, theta 2, etcetera, theta m from an available sample x 1, x 2, etcetera, x n. So, this is an observed sample from the sample we should be able to estimate the parameters theta 1, theta 2, etcetera, theta m.

The theta 1, theta 2, theta m which are set of parameters is in fact, a function of sample itself because you are estimating. I repeat that theta 1, theta 2, theta m etcetera, are the parameters of the population and you are estimating from the sample, the associated parameter estimates you are obtaining the associated parameter estimate theta i cap.

So, whenever we put a cap here it means that it is an estimate. So, we denote for example, theta I cap as an estimate of theta I whenever we are talking about the estimates the estimates will be a function of the sample itself and therefore, as you change the sample you get a different estimate and therefore, the theta I cap is a function of the sample and because the sample is a random sample theta I cap itself becomes a random variable; that means, you are estimating these parameter and these parameter estimates themselves become random variables with their own probability distributions and their own moments.

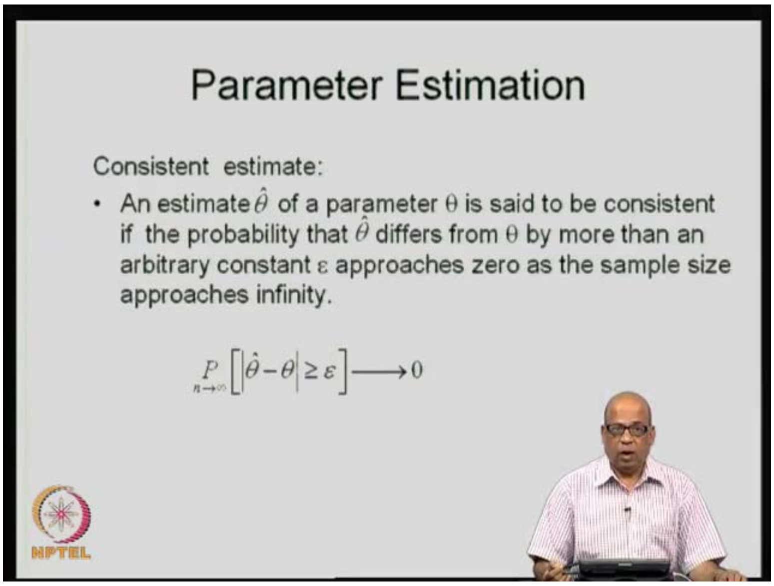
So, theta I cap has its own distribution and therefore, its own moments for example, it has its expected value, its own variance and so on. So, the question is what are the best estimates from the available samples sample x 1, x 2, etcetera, x n which in some sense is best for our purpose of using this particular distribution f of x? Now, we introduce two important properties of the estimators or estimates. So, we have one property called as unbiasedness of the estimate or we call them as unbiasedness estimate and we have the property of consistency of estimates or we call them the associated parameters estimates as consistent estimates of the parameter.

(Refer Slide Time: 36:07)



Let look at what we mean by the unbiasedness or unbiased estimates as I just mentioned or the theta cap which is a parameter is itself a random variable and the expected value of theta cap, if it approaches theta that is we say an estimate theta cap of a parameter theta is said to be unbiased if the expected value of theta cap is equal to theta itself and the bias if any is given by expected value of theta cap minus theta. So, what we are saying here is that your theta cap is such that your expected value of theta cap is equal to theta this does not mean that an individual theta cap that you get an individual estimate theta cap that you get will be exactly equal to theta or even close to theta it just means that the average of many independent estimates of theta will be equal to theta this is what is meant by a unbiased estimate.

(Refer Slide Time: 37:12)



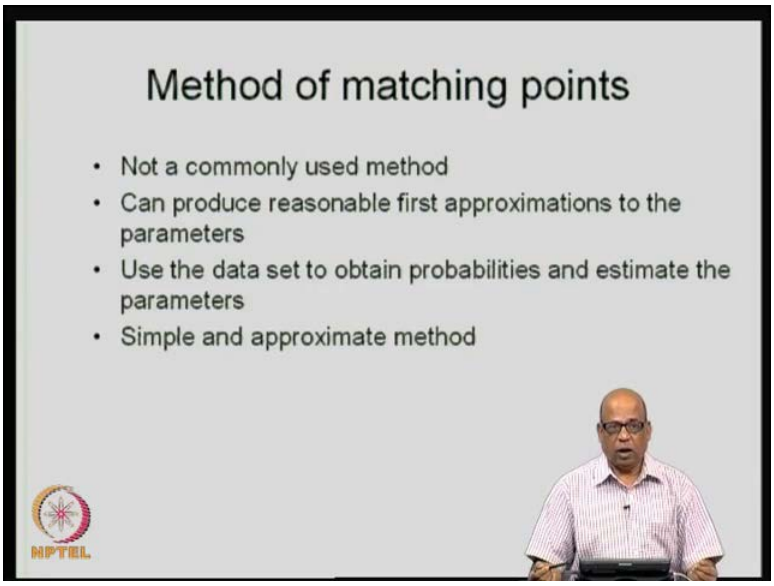
Similarly, we introduce a concept of a consistent estimate now an estimate theta cap of a parameter theta is said to be consistent if the probability that theta cap differs from theta by more than an arbitrary constant epsilon approaches 0 as the sample size approaches infinity. So, let say you have sample of 50 values and you estimate theta cap and then you have a sample of 100 values when you estimate theta cap and so on. So, as the sample size keeps on increasing if your probability of theta cap being close to theta with a distance of epsilon if this probability starts approaching 0 that is probability of theta cap minus theta the absolute value of this being greater than equal to an absolute arbitrary value epsilon if this approach is 0 as n tends to infinity.

N is a sample size as the sample approaches infinity then such a estimate is called as a consistent estimate. What we mean by this is that let us say we estimated a parameter theta cap with 50 values with some sample values and this probability is some this particular probability is some certain value and then as you increase your sample size. This probability becomes lower increase your sample size again the probability becomes much lower and so on. So, as the sample size increases the probability this probability starts becoming closer to 0 then it is called as a consistent estimate.

However in certain situations it may happen that as your sample size is increasing the estimates that you are obtaining will lead to these probabilities fluctuating around 0 not necessary approaching 0, sometimes they approach 0, sometimes they go away from 0, etcetera, they will become inconsistent estimates So, in certain situations will be interested in both unbiased as well as consistent estimates and therefore, it is important for us to understand this concept and test for unbiasedness as well as consistency of estimates. Now having introduced the concept of parameter estimate, we will now look at some methods which are available for estimating the parameter.

We will typically used the method of matching moments, method of matching points, method of moments and the method maximum likelihoods. We will introduce these three methods for estimating the parameters or there is also a method of graphical plotting method of potting which is a really approximate method which is used really as first cut rough estimate, but we will not deal with that particular method in this course. With the computers available etcetera, it is method of the graphically method is almost outdated now. In the methods of matching points what we do is essentially, we look at the samples values. From the sample values we estimate certain probabilities as relative frequency of such things and then we equate these probabilities with the probabilities as expected from the particular distribution and then from the associated probabilities we obtain the parameters.

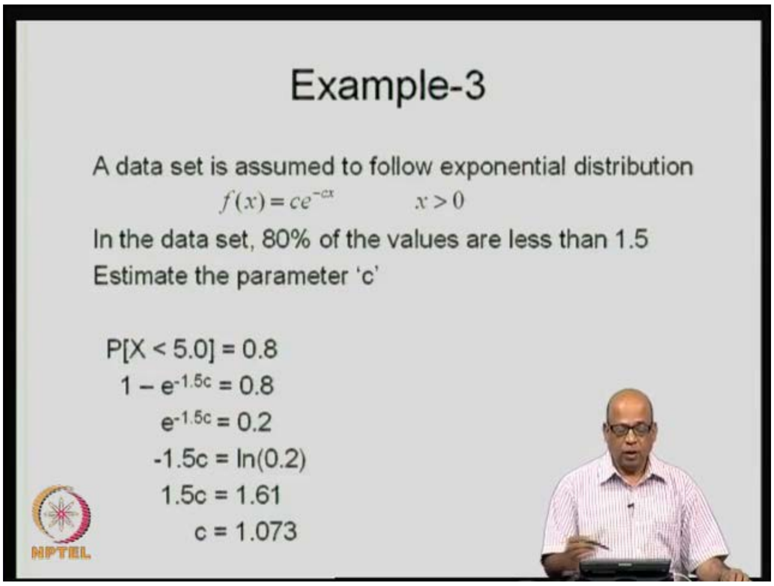
(Refer Slide Time: 40:40)



This is not a very commonly used method; however, when you have samples from which to we can estimate let us say probability of x being greater than equal to a given value is 80 percent and so on. When you can estimate those things as the first cut estimated of parameters you can use these. It is a simple and approximate method. Let us take an example here, a data set is assumed to follow the exponential distribution f of x is equal to c e to the power minus c x. Now from the data set that is available with us we see that the 80 percent of the values are less than 1.5. So, we use this fact to estimate the parameter c. So, what is it? That is given probability of x being less than equal to 5 in this particular case this one written as 1.5.

This is written as 5 here it is actually 1.5. So, probability of x being less than according 1.5 is given as 80 percent. So, here we are saying this should have been 1.0. So, probability of x being less than equal to 1.5 is 0.8 because 80 percent of the value are less than 1.5 then from this we write you know that f of x for there is a capital F of X for the exponential distribution recall that it is 1 minus e to the power minus lambda x in this particular case lambda is c.

(Refer Slide Time: 41:20)

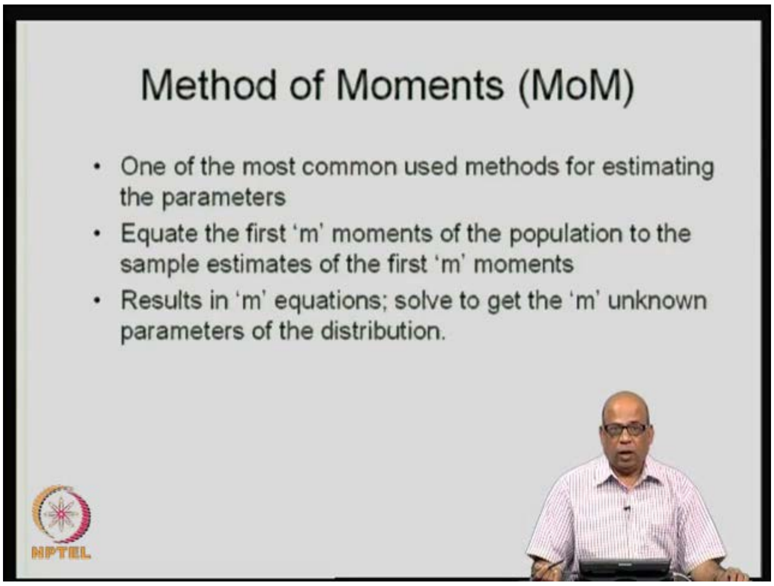


So, we write this as 1 minus e to the power minus c into x, x is 1.5, this equal to 0.8. So, e to the power 1 minus 1.5 c is equal to 0.2 from which you get c is equal to 1.073. So, essentially what we did here is that we looks at the sample we estimated some probability from the sample. So, we said than 80 percent of values are less than 1.5 we use that fact and then obtain c this is one way of estimating the parameters.

We just use one of the facts from the samples that 8 percent of the values are less than 1.5 and we obtain c this may be quite an approximate way of estimating the parameter. So, this may not be really the parameter estimate that will emerge from large sample values where we use the method of moments or method of maximum likelihood, which will introduce just now a better way of estimating a parameter and one of the most commonly used methods is by using the method of moments itself.

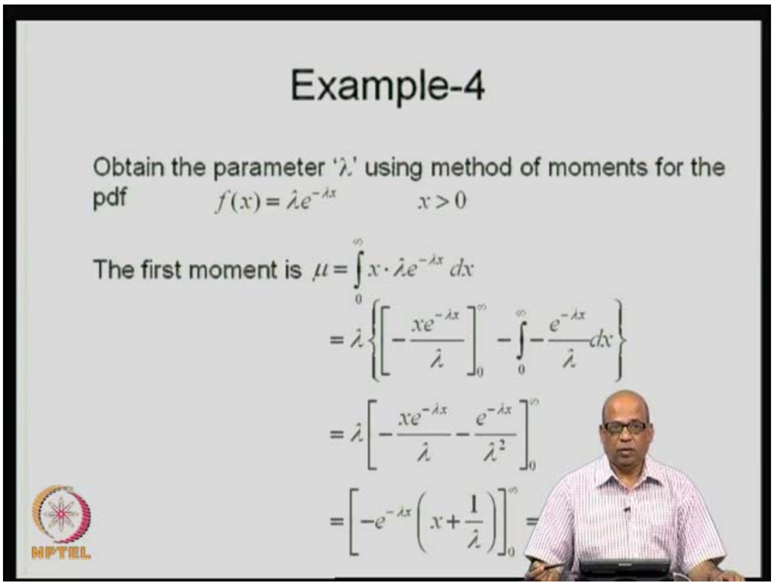
So, what we do here is that, let say a particular distribution has m number of parameters theta 1, theta 2, etcetera, m we equate the sample moment recall that the first sample moments is the sample estimate of the first moment is the x bar and similarly sample estimate of the second moments is x square or the variance estimates of the variance and so on. So, from the sample we can obtain the sample estimates of the moments first moments second moments third moment and so on. We equate these with the moments obtained on the population or the moments of the theoretical distribution and then equate them generate m number of equations by equating the m moments of the sample with the m moments of the population generate m number of equations solve these m equations to obtain the m parameters.

(Refer Slide Time: 44:36)



So, this actually results in m equation and we solve the m unknown parameters of the distribution. So, again we will take the example of the exponential distribution. So, we have a single parameter lambda here. So, we take the first moment. So, first moment of this distribution is mu is equal to 0 to infinity because x is greater than 0 here 0 to infinity x f of x d x, x into lambda e to the power minus lambda x d x. So, we integrate this by using method of a separation and then we get mu is equal to 1 by lambda. So, this is 0 to infinity essential what we are doing here is using the fact that we are integrating u v is equal to u v dash minus integral v u dash that is what you are using here and then you simplify that and get mu is equal to 1 over lambda for 0 to infinity, this is fairly straight forward you can refer to this later.

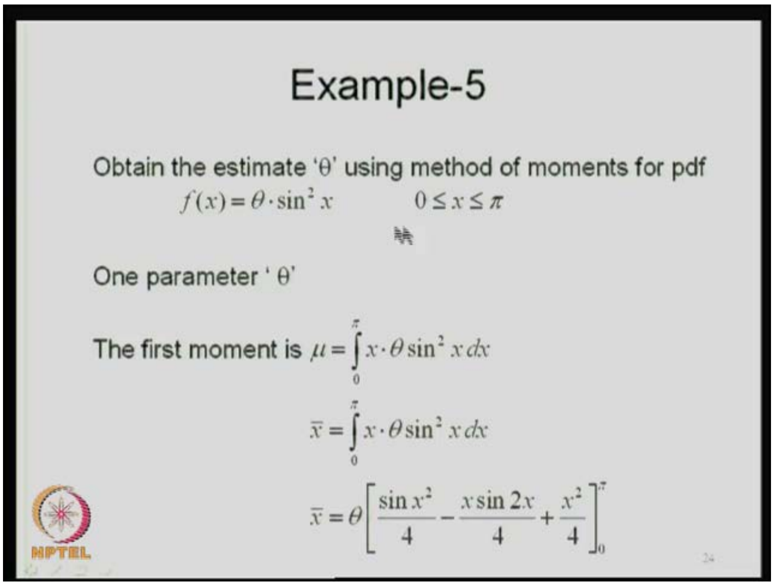
(Refer Slide Time: 45:19)



So, mu is equal to 1 over lambda now, what is sample estimate of mu? Mu is the first moments. So, if you have a sample how do we estimate mu from the sample we estimated by x bar which is the arithmetic average. So, we write this ass x bar is equal to 1 by lambda and therefore, lambda cap which is the sample estimate of a parameter lambda will be equal to 1 by x bar. So, this is how we proceed for a distribution having any number of parameters. If you have two parameters, what would do you will take two moments first two moments one is the mean and the second one is the second moments about the mean itself which is a variance and then you equate to the sample estimates of these corresponding moments and then obtain two equations solve for the two equations and then obtain the two parameters.

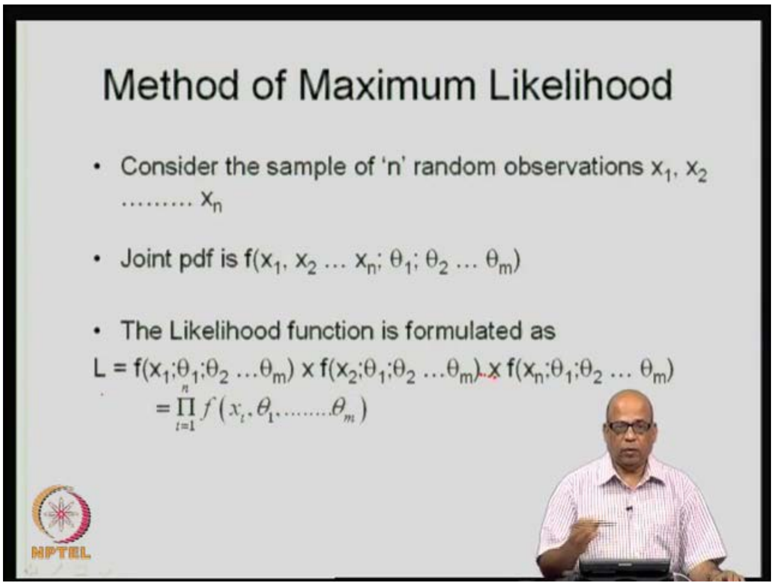
So, let us look at one other example here of which has again a single parameter theta. So, f of x is equal to theta sin square x for x varying between 0 and pi you can verify that this, in fact, is a valid pdf by integrating this between 0 and pi you will get one. So, this has one parameter theta. So, again we will take the first moment. The first moment of this would be 0 to pi x into f of x which is theta sin square x d x. Again we integrate by parts you get you can either use mat lab or you simply integrate in by pass you will get x bar is equal to theta sin x square by 4 minus x sin 2 x by 4 plus x square by 4 between 0 and pi.

(Refer Slide Time: 47:42)



So, that turns out to be x bar is equal to theta pi square by 4 remember here, we said x bar is an estimate for mu and therefore, we get theta cap here. So, theta cap will be 4 x bar by pi square. So, the method of moments we take those many moments as we have number of parameters and generate those many equations from these and equate and obtain the associated parameter. Now, we introduce a slightly more rigorous way of estimating the parameters which is the which is the method of maximum likelihood we have the sample consisting of n values x 1, x 2, etc,, x n. Now, your f of x if write it for x 1, x is equal to x 1, x is equal to x 2, etcetera, then we define a likelihood function as f of x 1, theta 1, theta 2, etcetera, theta m into f of x 2, theta 1, theta 2, etcetera, theta m into f of x n theta 1, theta 2, etcetera, theta m and so on. The basis of this is that is your write it for n terms here.

(Refer Slide Time: 51:06)

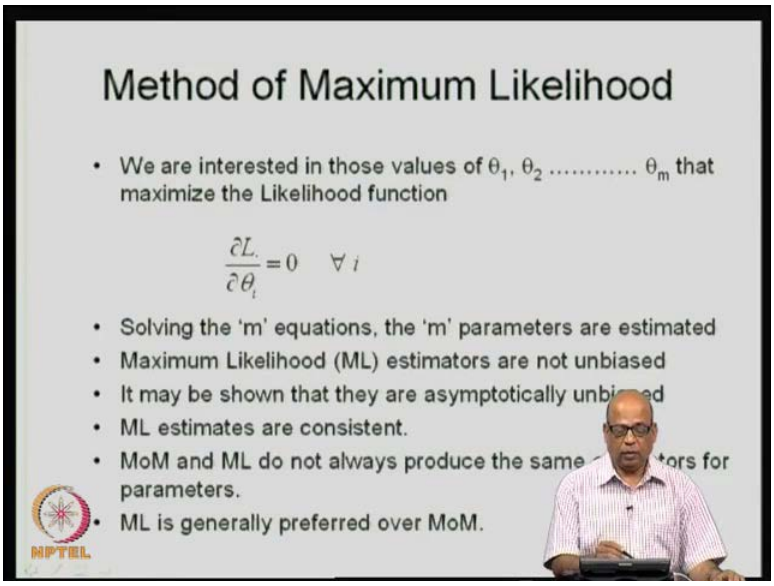


So, this is for n terms. The basis for this is that the once a sample has occurred x 1, x 2, etcetera are independence. So, we are looking for those set of parameter theta 1, theta 2 etcetera, theta m which will maximize the likelihood of this sample x 1, x 2 etcetera, x n appearing from that particular pdf. So, in a sense we are writing it as a joint distribution of f of x 1, f of x 2, f of x 3, etcetera and because they are independent x 1, x 2, etcetera, is the sample of which has actually occurred.

We obtain the likelihood function as product of f of x 1 into f of x 2 etcetera, for those parameters theta 1, theta 2, and etcetera. So, we first define the likelihood function as product of f of x i theta 1, theta 2, etcetera, over i is equal to 1, 2, n. This is the likelihood function and we are seeking those parameters theta 1, theta 2, etcetera, which will maximize the likelihood l which means we are looking for that set of parameter theta 1, theta 2, etcetera, theta m which results in the maximum likelihood of obtaining this particular sample x 1, x 2, etcetera, x n that is the principle here So, we formulate the likelihood function and then we maximize the likelihood function with respect to theta one theta two etcetera, theta m generate m number of equations by taking the first derivative, because we are looking at the maximization of L here maximization of L associated with the theta. So, with respect to each of the theta I, you differentiate that equated to 0 generate m number of equations and solve the m equations to get the parameters m parameters.

So, we just used method of moments as well as method of we have just introduce the method of maximum likelihood the maximum likelihood estimates are not unbiased estimates remember we also introduce earlier the concept of an unbiased estimate and the consistence estimate maximum likelihood estimates are not unbiased. However as you estimate using maximum likelihood method with several samples large number of samples then they maybe asymptotically unbiased. So, it may be shown that the maximum likelihood estimates are asymptotically unbiased.

(Refer Slide Time: 51:52)

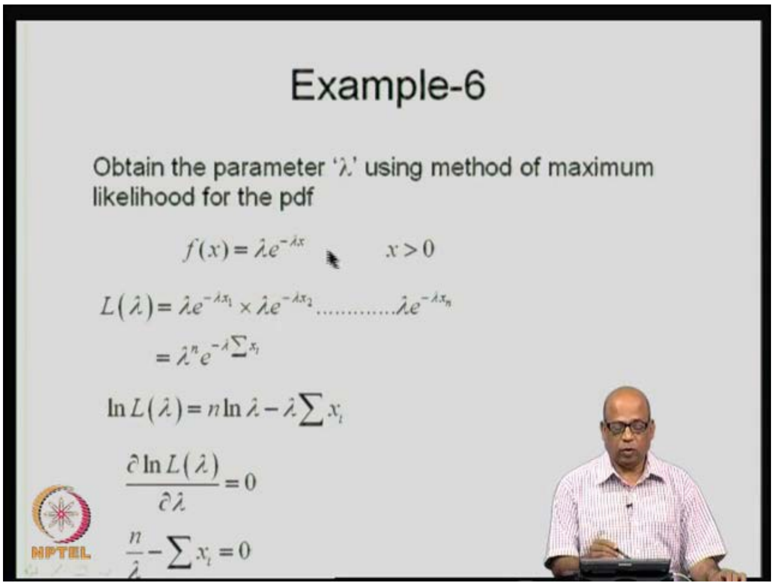


The maximum likelihood estimates are consistent. So, we just introduce the consistency concept. So, the maximum likelihood estimates are consistence method of moments and likelihood estimates method of maximum likelihood do not always produce a same estimate of false parameter. So, this must be remembered that is, you may use method of moments you may come up with some estimates for the parameters, but when we use the methods of maximum likelihood you may come up with some other estimate for the parameters. When this happens, the maximum likelihood is generally preferred over method of moments.

So, the maximum likelihood estimates are generally preferred over the methods of moments let us take a simple example to demonstrate how we estimate the parameters, using the methods of maximum likelihood. So, we will again take the same example as the exponential distributions as we consider for the method of moments f of x is equal to lambda e to the power of lambda x, the parameter is lambda here. We formulate the likelihood function l of lambda is equal to lambda e to the power minus lambda x 1 into lambda e to the power minus x 2 etcetera. So, essentially what we do is we formulate the likelihood function by putting x is equal to x 1 into x is equal to into f of x 2 which is by putting x is equal to x 2 and so on.

This is the product lambda e to the power minus lambda x 1 lambda e to the power minus lambda x 2 and So, on which is written as lambda to the power n multiplying it 10 times e to the power minus lambda sigma x I because this would be e to the power minus lambda x 1, plus x 2 plus x 3 etcetera, inside bracket. So, this is minus lambda e to the power.

(Refer Slide Time: 54.32)

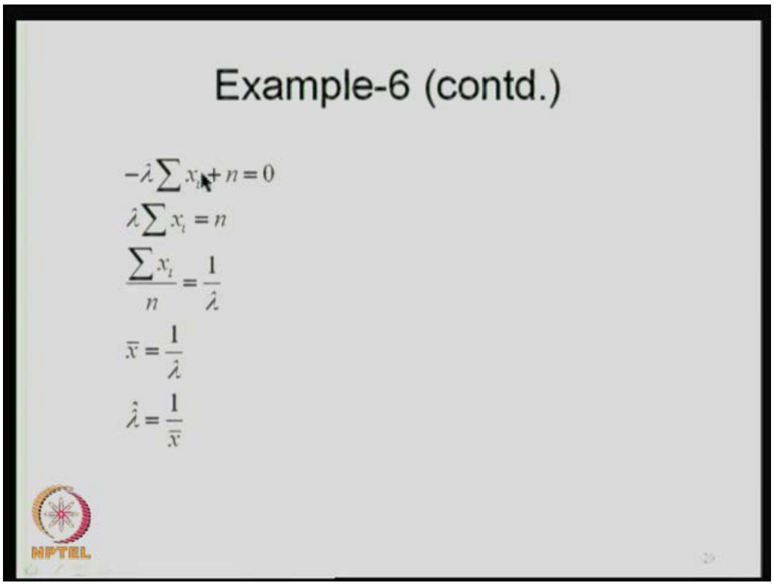


I will repeat, lambda to the power n e to the power minus lambda summation of x I over I is equal to 1 to n. It is often convenient to take the logarithm of these, logarithm of the likelihood function because log of x is a monotonic function of x. Therefore, log of x will have the maximum at the same place where x has the maximum ad therefore, if you take the logarithm of likelihood and the look at that value of lambda in that particular case at which the log L of m has the maximum the same value of lambda also corresponds to the place where L of lambda has maximum.

As we are looking for that particular value of lambda, it maximizes the likelihood function L of lambda. So, the log of log likelihood will be equal to n lambda n log of lambda minus lambda sigma of x I where simply taking the log of this we differentiate this with respect to lambda, because we are looking for the maximum value of log of lambda differentiate this with respect to lambda equated to 0. You get n over lambda minus sigma x I here n over lambda minus sigma x I your differentiating with respect to lambda will be equal to 0 and from this you get this is minus lambda sigma x I plus n is equal to 0 from this you get sigma x I by n equal to 1 over lambda.

What is sigma x y by n that is x bar. So, x bar is equal to 1 by lambda. So, we write from this lambda I cap which is an estimate will be equal to 1 by x bar. So, this is how we estimate parameters using the maximum likelihood method. Essentially we formulate the likelihood function which is simply the product of f of x at x is equal to x 1 into f of x at x is equal to x 2 and so on.

(Refer Slide Time: 56:04)



So, f of x I for I is equal to 1 to n, because you have the sample x 1, x 2, x 3, etcetera, x n we formulate the likelihood function then differentiate the likelihood with respect to the each of the parameters theta 1, theta 2, etcetera, theta m. Generate m equations equate it to 0 that is the first derivative equate it to 0, because you are looking for the maximum value the necessary condition is that the first derivative is equal to 0 generate m equations solve the m equations to get your m parameters. So, in this class today what we did is we went on to introduce the type III extreme value distributions and specifically we covered the Weibull distribution which is used generally for the al lower extreme. For example, the minimum values of stream flow and so on. We introduced the Weibull distribution and solved a couple of examples. Then we went on to discuss the parameter estimation.

We introduce the concept of an unbiased estimate and also a consistence estimate and we covered three methods of estimating the parameters, one is the method of matching parameters where we simply look at the sample and estimate certain parameters from the sample and we estimate certain probabilities, from the sample and equate it to the probabilities as obtained from the theoretical distribution. In the method of moments, we take those many moments as we have a number of parameters and thus generating those many equations and solving these equations we get parameters estimates of the parameters.

In the method of maximum likelihoods we define a likelihood function and obtain those parameters which will maximize this likelihood how do we do this, we take the first derivatives of the likelihood functions with respect to each of this parameters. Equate it to 0, because that is the necessary condition for the maximum of a function, equated it to 0 generate m number of equation when we have m number of parameters and then solve then to get the parameters. So, we continue this discussion in the next class thank you.